

⑬ 日本国特許庁 (JP)
公開特許公報 (A)

⑪ 特許出願公開
昭58-52696

⑤ Int. Cl.³
G 10 L 1/00

識別記号

庁内整理番号
7350-5D

⑬ 公開 昭和58年(1983)3月28日

発明の数 1
審査請求 未請求

(全 7 頁)

⑭ 音声認識装置

① 特 願 昭56-150752
② 出 願 昭56(1981)9月25日
③ 発 明 者 山邊巖
東京都江戸川区東小岩3-17-
2
④ 発 明 者 戸田明

東京都葛飾区西亀有4-14-6
発 明 者 善本正一
横浜市緑区桜台11-5
出 願 人 大日本印刷株式会社
東京都新宿区市谷加賀町1丁目
12番地
代 理 人 弁理士 武蔵次郎 外1名

明 細 書

1. 発明の名称

音声認識装置

2. 特許請求の範囲

(1) 音声信号から抽出した各音節ごとの特徴データを標準データと比較して音声認識を行なう方式の音声認識装置において、少くとも2種類の異なる特徴データ抽出方式により音声認識を行なつて複数の認識結果を得るようにした音声認識手段と、これら複数の認識結果を入力とする総合評価手段とを設け、複数の認識結果の総合評価による認識率の高い音声認識結果が得られるように構成したことを特徴とする音声認識装置。

(2) 特許請求の範囲第1項において、それぞれ専用の音声認識プロセッサを備え独立して音声認識処理動作が可能で少くとも2系統の音声認識手段を設け、異なる複数の特徴データ抽出方式による音声認識処理動作を単音節ごとに同時に開始して上記複数の認識結果を得るように構成したことを特徴とする音声認識装置。

(3) 特許請求の範囲第1項又は第2項において、上記複数の認識結果のそれぞれに対応し、それぞれの認識結果の正誤に応じて順次更新されてゆく複数の信頼度係数を算出する手段を設け、上記評価手段を、これら複数の信頼度係数のそれぞれに対応する上記複数の認識結果のそれぞれに適用した上で総合的に評価する手段で構成したことを特徴とする音声認識装置。

3. 発明の詳細な説明

本発明は、人間の音声によつて入力された言語データを単音節ごとに判断して対応する文字データを得るための音声認識装置に関する。

このような音声認識装置としては、従来からコンピュータを用い、入力された音声から各音節ごとの特徴データを抽出し、これをあらかじめ用意してあるそれぞれの音節ごとの標準特徴データと比較して音声の認識を行なう装置が主として採用されていた。

そして、このときの特徴データの抽出方式としては、例えば以下に示すような種々の方式のもの

BEST AVAILABLE COPY

が使用されていた。

1. 線形予測係数又はPARCOR係数方式。
2. 零交差波解析方式。
3. 周波数スペクトラム方式。
4. 第1フォルマント抽出方式。
5. 無音期間の後に続く高周波成分による方式。
6. 振幅による方式。

ところで、これらの方式を用いて音声の特徴データの抽出を行ない、音声認識を行なつてみると、各方式には音節の種類により得手、不得手があり、単音節を構成する音素の特徴を良く現わす方式とそうでない方式とに別れ、音声認識結果の認識率にも音節の種類によつて差があることが判る。

そして、一般に、日本語を対象とした場合、音声の中の音素とその特徴データを良好に抽出する方式との対応は、例えば第1表に示すようになつてゐる。

第1表

i, e, a, o, u, m, n, N	線形予測係数、PARCOR係数
h	零交差波解析
r	振幅、周波数スペクトラム、零交差波解析
p, t, k	振幅、第1フォルマント、零交差波解析
	無音区間の後に続く高周波成分

従つて、従来の音声認識装置においては、音節の種類によつては充分な認識結果が得られないという欠点があつた。

本発明の目的は、上記した従来技術の欠点を除き、必要とする音節のすべてにわたつて充分に高い認識結果が得られるようにした音声認識装置を提供するにある。

この目的を達成するため、本発明は、入力された音声の各音節ごとに複数の異なつた方式による特徴データの抽出を行ない、これによつて得た複

数の認識結果の総合的な評価により装置全体としての認識結果を得るようにし、それぞれの特徴データ抽出方式ごとの得手、不得手を相互に補い、或る方式による認識結果がその方式において得手とする音節に対するものであつたときには、その認識結果を高い信頼度のもとで評価し、反対に不得手を音節に対するものであつたときには低い信頼度のもとで評価するようにした点を特徴とする。

以下、本発明による音声認識装置を実施例によつて説明する。

上記したように、本発明においては、複数の異なつた特徴データ抽出方式による音声認識を入力される音声の単音節ごとに行なう必要があり、これを1台のプロセッサで行なわせようとするれば認識時間が長時間化し、入力音声の実時間処理が困難になり、さらに処理動作が極めて高速度の高価なプロセッサと入力音声を書きこくための記憶装置などが必要になるため、著しくコストアップを招きやすい。

そこで、以下に示す本発明の実施例においては、

それぞれの異なつた特徴データ抽出方式による音声認識処理装置をそれぞれ専用に設け、各方式による認識処理を並列に実行させることにより、比較的低速のプロセッサ、例えばマイクロプロセッサを用いても充分に短い時間で認識処理が行なえるようにしてコストダウンを可能にしている。また、この実施例においては、上記した複数の音声認識処理装置によつて得た複数の認識結果を総合的に評価し、本発明による音声認識装置全体としての最終的な認識結果を得るための処理装置にも専用のプロセッサが用いられており、いわゆるマルチプロセッサ構成となつてゐる。そして、この結果、各プロセッサに必要なソフトウェアの単純化が可能になつてゐる。

さて、第1図は本発明の一実施例を示すブロック図で、1はマイクロホンと増幅器などからなる音声信号入力部、2は線形予測係数方式（又はPARCOR係数方式）の認識プロセッサ、3は零交差波分析方式の認識プロセッサ、4は周波数スペクトラム方式の認識プロセッサ、5は第1フォル

マント抽出方式の認識プロセッサ、6は振幅方式の認識プロセッサ、7は無音期間に続く高周波成分方式の認識プロセッサ、8は評価プロセッサである。

各方式の認識プロセッサ2～7は音声信号入力部1からの信号を共通に取入れ、認識結果をそれぞれ評価プロセッサ8に通知する。

評価プロセッサ8は各認識プロセッサ2～7からの認識結果をそれぞれ評価し、この装置全体としての認識結果（以下、これを最終認識結果という）を算出する。また、この評価プロセッサ8は、それぞれの認識プロセッサ2～7の動作を制御し、さらに外部装置との情報交換も受持っている。

従つて、音声信号入力部1と各認識プロセッサ2～7との間はアナログ信号回路で結ばれ、これら認識プロセッサ2～7と評価プロセッサ8との間はデジタル通信回路で結ばれている。

次に、第2図はそれぞれの認識プロセッサ2～7の一実施例で、10はそれぞれの特徴データ抽出方式による特徴抽出用ハードウェア、11はマ

従つて、この第2図に示した装置は内部バス16で結合されて全体で第1図に示したそれぞれの認識プロセッサ2～7を形成し、音声信号が入力されるとその各音節ごとにそれぞれの方式にしたがつた特徴データの抽出を行ない、それを辞書メモリ13から読出した標準データと比較して認識結果を得るようにCPU11による処理が行なわれることになる。

また、第3図は第1図における評価プロセッサ8の一実施例で、20は通信回路、21はマイクロプロセッサ（CPU）、22は作業用メモリ、23は信頼度管理テーブル、24はプログラムメモリ、25は通信回路、26は内部バスである。

通信回路20は第1図及び第2図に示したそれぞれの認識プロセッサ2～7との間での情報交換を行なうハードウェアであり、同じく通信回路25は外部装置、例えばモニタを含むプリンタ、電子複写装置などとの間での情報交換を行なうハードウェアである。

信頼度管理テーブル23は各認識プロセッサ2

イクロプロセッサ（CPUという）、12は作業用のメモリ、13は音声辞書メモリ、14はプログラムメモリ、15は通信回路、16は内部バス（BUSという）である。

ハードウェア10はそれぞれの特徴データ抽出方式に応じて構成され、入力されたアナログ音声信号をその方式にしたがつて特徴抽出処理してからデジタル化して出力する。

CPU11はプログラムメモリ14に格納されている音声認識プログラムにしたがつて所定の処理を行なう。

音声辞書メモリ13には入力される音声信号として予定されている各音節の音声見本より得た各音節に対応する特徴データ、即ち、標準データが格納されている。

通信回路15は第1図に示した評価プロセッサ8との間で情報交換を行なうためのハードウェアである。なお、作業用メモリ12、内部バス16などは通常のマイクロコンピュータと同じであるから説明は省略する。

～7の各音節ごとの認識結果についての信頼度情報を記録しておくメモリである。

なお、CPU21、作業用メモリ22、プログラムメモリ24、内部バス26などは第2図と同じである。

従つて、この第3図に示した装置は内部バス26で結合されて全体で第1図に示した評価プロセッサ8を形成し、各認識プロセッサ2～7の制御と、それから得られた各音節ごとの認識結果の総合的な評価とを行ない最終認識結果を外部装置に出力すると共に、この最終認識結果に対する外部装置からのフィードバック情報を受け入れ、信頼度管理テーブル23の内容を更新する働きをする。

また、以上の説明から明らかなように、第1図に示した本発明の実施例は、全体として複数の認識プロセッサ2～7と評価プロセッサ8とを通信回路によつて結合した、いわゆる疎結合マルチプロセッサを形成している。

次に動作について説明する。

既に説明したように、この実施例による音声認識装置全体の動作、及び外部装置との間の情報交換は全て評価プロセッサ8によつて行なわれる。

そして、各認識プロセッサ2～7は評価プロセッサ8からの制御命令により動作し、その結果である認識結果を評価プロセッサ8に送り出す。

このとき、評価プロセッサ8から各認識プロセッサ2～7のそれぞれに送られる制御命令とデータは、

1. 音声辞書格納命令、及び音声辞書データ、
2. 音声辞書作成（更新）命令、及び作成（更新）される音節名、
3. 音声認識作業開始命令、
4. 音声辞書転送命令、

であり、これに対して各認識プロセッサ2～7のそれぞれから評価プロセッサ8に送られる情報は、上記1～4に対応して

1. 音声辞書格納完了信号、
2. 音声辞書作成（更新）終了信号、
3. 音声認識結果、

13

ら各認識プロセッサ2～7に送り出す代りに、音声辞書作成（更新）命令と作成（更新）される音節名を送り出し、一方、各認識プロセッサ2～7は音声辞書格納完了信号の代りに音声辞書作成（更新）終了信号を評価プロセッサ8に送り返す。

なお、この音声辞書作成（更新）作業については周知のとおりに行なえばよいので、説明は省略する。

さて、こうして音声認識処理が開始すると評価プロセッサ8には各認識プロセッサ2～7から6種類の認識結果が送られてくる。このとき、各認識プロセッサ2～7から送られてくる認識結果は、日本語を対象とした場合、その音節A～Nの68種類（だく音、半だく音を含めれば101種類）のうちのいずれか1音節に対応するものとなつてゐる。

そこで、評価プロセッサ8は各認識プロセッサ2～7のそれぞれの認識結果（音節）からその音節に対応する信頼度 r を信頼度管理テーブル23から取り出す。

4. 音声辞書データ、

となつてゐる。

そこで、まず、装置が動作可能な状態に操作されると、評価プロセッサ8は音声辞書格納命令とそれに続いて外部の所定のメモリに収容してある音声辞書データを各認識プロセッサ2～7に送出し、各認識プロセッサ2～7の辞書メモリ13に認識処理に必要な標準データを格納する。そして、辞書メモリ13に対する標準データの格納が完了すると、対応する認識プロセッサ2～7から音声辞書格納完了信号が評価プロセッサ8に送られる。

そこで、次に、評価プロセッサ8は各認識プロセッサ2～7に音声認識作業開始命令を送り、これにより各認識プロセッサ2～7はそれぞれの特徴データ抽出方式による音声認識作業を開始し、音声が入力されるとそれぞれの方式による認識結果を得、それらを評価プロセッサ8に転送する。

一方、オペレータが突つたりして標準データの作成（更新）が必要になつたときには、音声辞書格納命令と音声辞書データを評価プロセッサ8か

14

ここで、信頼度管理テーブル23に格納されている信頼度を R とすれば、

$$(R | r(1, 1), r(1, 2), r(1, 3) \dots, r(a, 1))$$

a = 日本語の音節数（1～68となる）、

i = 認識プロセッサの数（この実施例では6である）。

となり、 R の要素数は $(a \times i) = (68 \times 6)$ となる。

そして、 $r(j, k)$ は第 k 番目の認識プロセッサの音節を j としたときの認識結果が正しいと思われる信頼度を表わし、

$$0 \leq r(j, k) \leq 1$$

であり、この値が1.0に近いほどその認識結果に対する信頼度が高い。

そこで、各認識プロセッサからの認識結果に対する信頼度を R_i とすれば、

$$(R_i | r_1(a), r_2(a), r_3(a) \dots, r_i(a))$$

となり、 $r_i(a)$ は第 i 番目の認識プロセッサによる認識結果の音節 a に対する信頼度を表わし、信

信頼度 R_i の要素数は 1 個となり、第 1 図の実施例では 6 個となる。

また、信頼度 r_i の属性となる音節の種類は、全ての認識プロセッサが同じ音節と認識した場合に、全ての a の値は同じであるので 1 種類であり、全ての異なつた音節であると認識したときには 1 種類になる。

次に、各認識プロセッサ単位の信頼度 R_i を音節の種類ごとの信頼度 $S_{a\ell}$ に変換する。ここで $S_{a\ell}$ は、

$$(S_{a\ell} | S_{a1}, S_{a2}, S_{a3}, \dots, S_{a\ell})$$

但し、 ℓ は信頼度 R_i 中に含まれる音節の種類である。

であり、この変化は次のようにして行なう。

$$S_{a\ell} = [R_i = a | S_{a\ell} = 1 - \pi \{1 - r_i(a)\}]$$

なお、 π は積和を渡す。

そして、信頼度 R_i 中 a という音節を認識結果に持つ全ての $r_i(a)$ に対して、 $1 - \pi \{1 - r_i(a)\}$ を計算したものを音節 a に対する信頼度 S_a とする。この信頼度 S_a は信頼度 R_i に含まれる音節

いる信頼度とを外部のメモリに待避させるように動作する。

そして、このうち音声辞書の特選は評価プロセッサ 8 から各認識プロセッサ 2～7 に音声辞書転送命令を送り、これに応じて各認識プロセッサ 2～7 はその音声辞書メモリ 13 から音声辞書データを読出して転送することにより行なわれる。

以上説明したように、この実施例によれば、入力音声信号の各音節ごとの認識処理をそれぞれ異なつた特徴データ抽出方式による専用の認識プロセッサ 2～7 で同時に開始し、並列に行なっているから、比較的低速の CPU で構成しても処理時間が永くなる虞れが少なく、しかも、複数の異なつた特徴データ抽出方式による認識結果をそれぞれの方式によつて決まる信頼度を加味して総合的に評価することにより最終認識結果を得ているから、特徴データ抽出方式の相違による音節ごとの得手、不得手が相互に補われ、充分に高い認識率をもつた結果を容易に得ることができる。

ところで、上記実施例においては、評価プロセ

の種類 ℓ 個存在するから、これに始めから番号を付して信頼度 $S_{a\ell}$ としてある。

そこで、評価プロセッサ 8 はこの信頼度 $S_{a\ell}$ を用い、認識プロセッサ 2～7 から得た認識結果の評価を行なつて最終認識結果を算出する。即ち、最も信頼度の高い認識結果は、 $S_{a\ell}$ 中で最大の信頼度 $S_{\max a} = \text{MAX}(S_{a\ell})$ を持つ音節 a となるから、認識プロセッサ 8 はこの信頼度 $S_{\max a}$ を持つ音節を最終認識結果として出力する。

こうして、入力された音声信号の一音節に対する認識処理を完了したら、ついで評価プロセッサ 8 は再び音声認識作業開始命令を各認識プロセッサ 2～7 に送り、次の音節の音声信号に対する認識処理に入つて動作を続行する。

次に、評価プロセッサ 8 は音声を入力するオペレータが変つたり、この音声認識装置が動作を停止させようとしたときには、各認識プロセッサ 2～7 のそれぞれの音声辞書メモリ 13 に格納されている音声辞書と、評価プロセッサ 8 の中に設けられている信頼度管理テーブル 23 に格納されて

評価プロセッサ 8 の信頼度管理テーブル 23 に最初に格納すべき信頼度データについては、特に説明しなかつたが、これは各特徴データ抽出方式による認識装置のそれぞれの音節に対する認識結果の過去の経験や実績に基づいて定めればよい。

しかしながら、このような音声認識装置においては、認識結果をモニタしてオペレータが判断し、誤つた結果が得られたときには訂正しながら使用されるのが通例であり、このような場合には認識結果についての正誤情報を得ることができる。

そこで、このようなときには、認識結果の正誤情報により信頼度管理テーブル 23 に格納すべき信頼度データを正誤情報に基づいて更新し、装置の使用を続行するにつれてさらに高い認識率が得られるようにすることができ、以下、この点についての一実施例について説明する。

まず、上記した信頼度 R は、原理的に次のように定義できる。

$$r(a, i) = \frac{n_c(a, i)}{n(a, i)} \quad \dots \dots \dots (1)$$

$n(a, 1)$ …… 第1番認識プロセッサが
入力音声 a と認識した
回数。

$nc(a, 1)$ …… 第1番認識プロセッサが
入力音声 a と認識し、
それが正しかつた回数。

そして、信頼度 R の更新は次のようにして行な
うようにする。

1. 第1番認識プロセッサの認識結果が、外部
からの正誤情報により正しかつたと判断され
たときには(1)式の分母と分子に共に1を加え

$$r(a, 1) = \frac{nc(a, 1) + 1}{n(a, 1) + 1} \quad \dots \dots (2)$$

とする。

2. 同じく、正しくなかつたと判断されたとき
には(1)式の分母にだけ1を加え

$$r(a, 1) = \frac{nc(a, 1)}{n(a, 1) + 1} \quad \dots \dots (3)$$

とする。

しかしながら、この更新方法では、 n と nc の

21

り。

$$r'(a, 1) = \frac{m \times r(a, 1) - 1 + \beta}{m}$$

ここで、 β は第1番認識プロセッサの今回の認
識結果 a が正しかつたとき1に、そして誤まつて
いたときには0になる数である。

また、 $r(a, 1) < rand(x)$ となつたときには
更新した結果、 $(m+1)$ 回前となつてしまふ認識
結果は限りであつたものとみなし、

$$r'(a, 1) = \frac{m \times r(a, 1) + \beta}{m}$$

によつて更新を行なうようにする。

こうして信頼度 $r(a, 1)$ の更新を全ての認識プ
ロセッサ2～6の信頼度管理テーブル23に格納
してあるデータについて行なえば、信頼度 R の更
新を容易に行なうことができ、上記実施例の音声
認識装置における最終認識結果の認識率をその使
用回数の増加とともにさらに改善することができ
る。

以上説明したように、本発明によれば、複数の

値を順次そのまま保存しておかなくてはならず、
しかも、これら n 、 nc の値は装置の使用回数に
伴つて無限なく大きくなるから処理に具合が悪
い。

そこで、この実施例では、現時点での信頼度 R
だけから更新後の信頼度 R' を算出するようにした。

まず、現時点での信頼度 R が現時点より m 回前
までの認識結果からの信頼度を反映していると假
定すれば、

$$n(a, 1) = m$$

$$nc(a, 1) = m \times r(a, 1)$$

となる。

次に更新後の信頼度 R' に更新したことにより
 $(m+1)$ 回前になつてしまふ認識結果が正しい答
であつたか誤りであつたかを信頼度 R の値から以
下のようにして決定する。

まず、 $rand(x)$ を0.0～1.0の値をとる一様
乱数とし、もしも、 $r(a, 1) \geq rand(x)$ ならば
更新した結果 $(m+1)$ 回前になつてしまふ結果を
正しいものとみなし、次の式によつて更新を行な

22

異なつた音声特徴データ抽出方法による認識結果
を総合的に評価して最終的な認識結果を得るよう
にしたから、それぞれの抽出方式における音節ご
との得手、不得手が相互に補われて充分に正しい
認識結果が得られることになり、従来技術の欠点
を除き、入力された音声において必要とする音節
の全てについて常に充分に高い認識率を保つて動
作することの可能な音声認識装置を提供すること
ができる。

4. 図面の簡単な説明

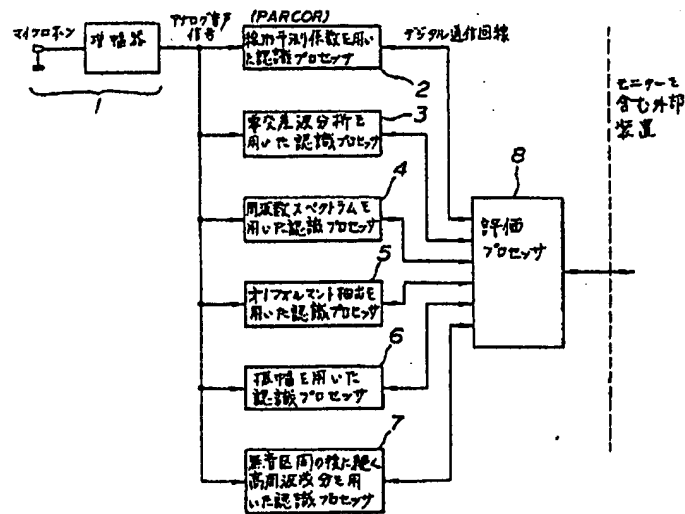
第1図は本発明による音声認識装置の一実施例
を示すブロック図、第2図は認識プロセッサの一
実施例を示すブロック図、第3図は評価プロセッ
サの一実施例を示すブロック図である。

1 …… 音声信号入力部、2～7 …… それぞれ異
なつた特徴抽出方式による認識プロセッサ、8 ……
評価プロセッサ。

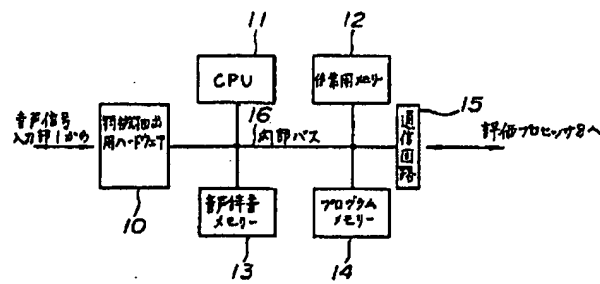
代理人 弁理士 武 副次郎 (ほか1名)



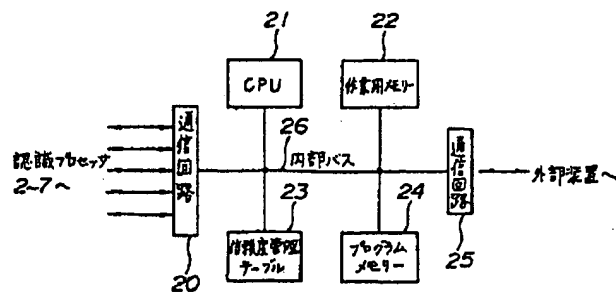
第 1 図



第 2 図



第 3 図



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☒ **OTHER:** copied, documents

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.